Microarray Bioinformatics

DOV STEKEL
Oxford University and Bius

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2003

Printed in the United States of America

Typeface Utopia 9.5/13.5 pt. and ITC Symbol *System* LaTeX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Stekel, Dov, 1971–

Microarray bioinformatics / Dov Stekel.

p. cm.

Includes bibliographical references and index.

ISBN 0-521-81982-2 (hbk.) – ISBN 0-521-52587-X (pbk.)

1. DNA microarrays – Mathematics. 2. DNA microarrays – Data processing. 3. DNA microarrays – Statistical methods. 4. Bioinformatics. I. Title.

QP624.5.D726S74 2003
572.8′636 – dc21 2003043959

ISBN 0 521 81982 2 hardback
ISBN 0 521 52587 X paperback

Contents

CHAPTER ONE

Microarrays: Making Them and Using Them

SECTION 1.1 INTRODUCTION

A DNA microarray consists of a solid surface, usually a microscope slide,[1] onto which DNA molecules have been chemically bonded. The purpose of a microarray is to detect the presence and abundance of labelled nucleic acids in a biological sample, which will hybridise to the DNA on the array via Watson–Crick duplex formation, and which can be detected via the label. In the majority of microarray experiments, the labelled nucleic acids are derived from the mRNA of a sample or tissue, and so the microarray measures gene expression. The power of a microarray is that there may be many thousands of different DNA molecules bonded to an array, and so it is possible to measure the expression of many thousands of genes simultaneously.

This book is about the bioinformatics of DNA microarrays: the mathematics, statistics and computing you will need to design microarray experiments; to acquire, analyse and store your data; and to share your results with other scientists. One of the features of microarray technology is the level of bioinformatics required: it is not possible to perform a meaningful microarray experiment without bioinformatics involvement at every stage.

However, this chapter is different from the remainder of the book. While the other chapters discuss bioinformatics, the aim of this chapter is to set out the basics of the chemistry and biology of microarray technology. It is hoped that someone new to the technology will be able to read this chapter and gain an understanding of the laboratory process and how it impacts the quality of the data. The chapter is arranged into two further sections:

Section 1.2: Making Microarrays, describes the main technologies by which microarrays are manufactured.
Section 1.3: Using Microarrays, describes what happens in a microarray laboratory when a microarray experiment is performed.

SECTION 1.2 MAKING MICROARRAYS

There are two main technologies for making microarrays: **robotic spotting** and **in-situ synthesis**.

[1] Historically, microarrays have also been produced using nylon filters and larger glass slides.

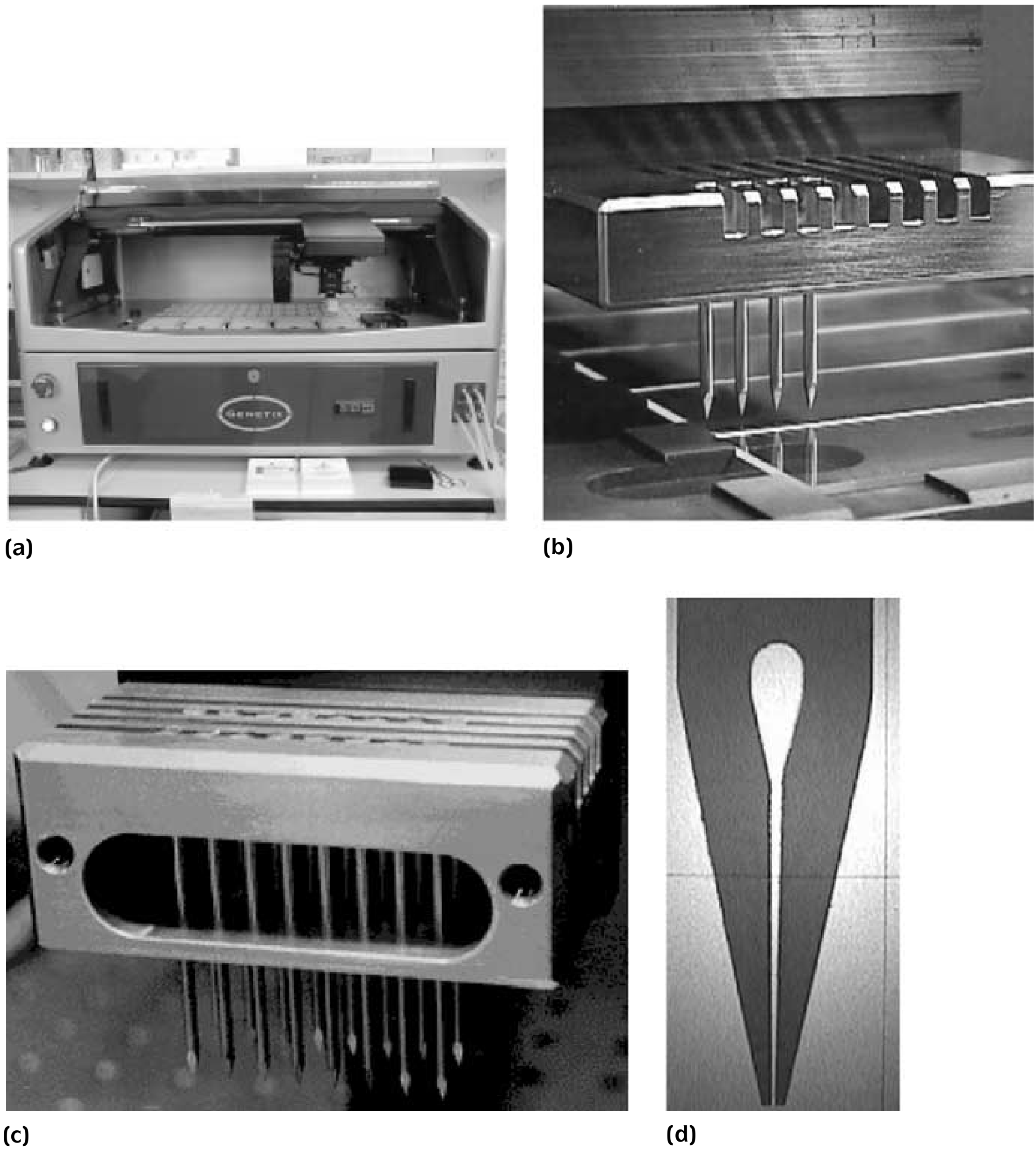


Figure 1.1: Spotting robot. (a) An example of a spotting robot. There are many different robots on the market; this is a Genetix spotting robot located at the Mouse Genetics Unit in Harwell, Oxfordshire. **(b)** The pins are held in a cassette in a rectangular grid, which in turn is held on a robot arm that can be moved between the microtiter well plates and the glass arrays to deposit liquid. **(c)** The number of pins in the cassette can vary. The more pins, the greater the throughput of the robot, but the greater the propensity for pin-to-pin variability. Each pin will spot a different grid on the array (Chapter 4). **(d)** Most pins in modern use have a reservoir that holds sample and so can print multiple features – usually on different arrays – from a single visit to the well containing probe. Earlier robots use solid pins, which can only print one feature before needing to collect more DNA from the well.

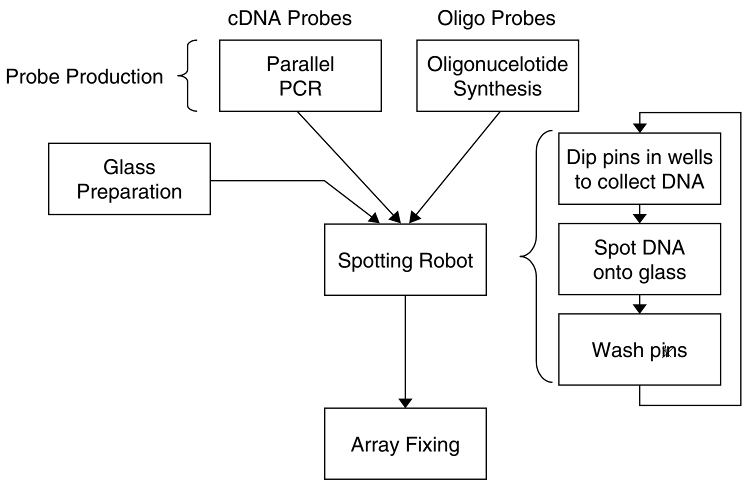


Figure 1.2: Spotted array synthesis. There are several steps involved in the synthesis of spotted arrays. First is the production of the probes. cDNA probes are made via highly parallel PCR; oligonucleotide probes have to be presynthesised. The next step is the spotting step. The robot arm moves the cassette containing the pins over one of the microtiter plates containing probe and dips the pins into the wells to collect DNA. The arm is then moved over the first array and the cassette is moved down so that the pins touch the glass and DNA is deposited on the surface. If more than one array is being synthesised, the cassette is moved to the subsequent arrays. Before collecting the next DNA to be spotted, the pins are washed to ensure no contamination. The final step of array production is fixing, in which the surface of the glass is modified so that no additional DNA can stick to it.

Spotted Microarrays

This is the technology by which the first microarrays were manufactured. The array is made using a spotting robot (Figure 1.1a) via three main steps (Figure 1.2):

1. Making the DNA probes[2] to put on the array
2. Spotting the DNA onto the glass surface of the array with the spotting robot
3. Postspotting processing of the glass slide

There are three main types of spotted array (Table 1.1), which can be subdivided in two ways: by the type of DNA probe, or by the attachment chemistry of the probe to the glass.

The DNA probes used on a spotted array can either be polymerase chain reaction (PCR) products or oligonucleotides. In the first case, highly parallel PCR is used to amplify DNA from a clone library, and the amplified DNA is purified. In the second case, DNA oligonucleotides are presynthesised for use on the array.

[2] There are now three camps in the microarray community as to what to call the DNA on the array and the DNA in solution. Throughout this book, we will use the "Southern" terminology and refer to the DNA on the array as *probes* and the labeled DNA in solution as *target*. Other researchers refer to the DNA on the array as *target* and the labeled DNA in solution as *probe*. More recently, MIAME (Minimal Information About a Microarray Experiment) introduced a new convention of referring to the DNA on the array as *reporters* and the DNA in solution as the *hybridisation extract*. MIAME conventions are described in full in Chapter 11, and MIAME terms are detailed in the Appendix.

TABLE 1.1

	Surface Chemistry	
DNA Probes	**Covalent**	**Non-covalent**
Oligonucleotides	✓	
cDNAs	✓	✓

Note: There are three types of spotted microarrays, which can be thought about in two different ways. The DNA probes can be oligonucleotides or cDNAs; the surface chemistry can be covalent or non-covalent. Oligonucleotide probes can only be attached covalently; cDNA probes can be attached either covalently or non-covalently. Covalent attachment is via an aliphatic amine (NH_2) group added to the 5′ end of the DNA probe, and consequently the probes are tethered to the glass from the 5′ end. Non-covalent attachment is via electrostatic attraction between amine groups on the glass slide and the phosphate groups on the DNA probe backbone; thus the DNA probe is attached to the glass by its backbone.

The attachment chemistry can either be covalent or non-covalent. With covalent attachment, a primary aliphatic amine (NH_2) group is added to the DNA probe and the probe is attached to the glass by making a covalent bond between this group and chemical linkers on the glass. With oligonucleotide probes, the amine group can be added to either end of the oligonucleotide during synthesis, although it is more usual to add it to the 5′ end of the oligonucleotide. With cDNA probes, the amine group is added to the 5′ end of the PCR primer from which the probes are made. Thus the cDNA probes are always attached from the 5′ end.

With non-covalent attachment, the bonding of the probe to the array is via electrostatic attraction between the phosphate backbone of the DNA probe and NH_2 groups attached to the surface of the glass. The interaction takes place at several locations along the DNA backbone, so that the probe is tethered to the glass at many points. Because most oligonucleotide probes are shorter than cDNAs, these interactions are not strong enough to anchor oligonucleotide probes to glass. Therefore, non-covalent attachment is usually only used for cDNA microarrays.

The DNA probes are organised in microtiter well plates, typically 384 well plates. Most modern spotting robots will use a number of plates to print arrays, so the plates are arranged in a "hotel," whereby the robot is able to gain successive access to each of the plates. The spotting robot itself consists of a series of pins arranged as a grid and held in a cassette (Figures 1.1b and 1.1c). The pins are used to transfer liquid from the microtiter plates to the glass array.[3]

There are a number of different designs of pins. The first spotting robots used solid pins (Figure 1.1b); these can only hold enough liquid for one spot on the array, thus requiring the pin cassette to return to the plate containing probe before printing the next spot. Most array-making robots today have pins with a reservoir that holds

[3] Not every spotting robot is a pin-based system: Perkin Elmer sell some robots which use a piezo-electric system to fire tiny drops of liquid onto the arrays. These are in the minority in the microarray field.

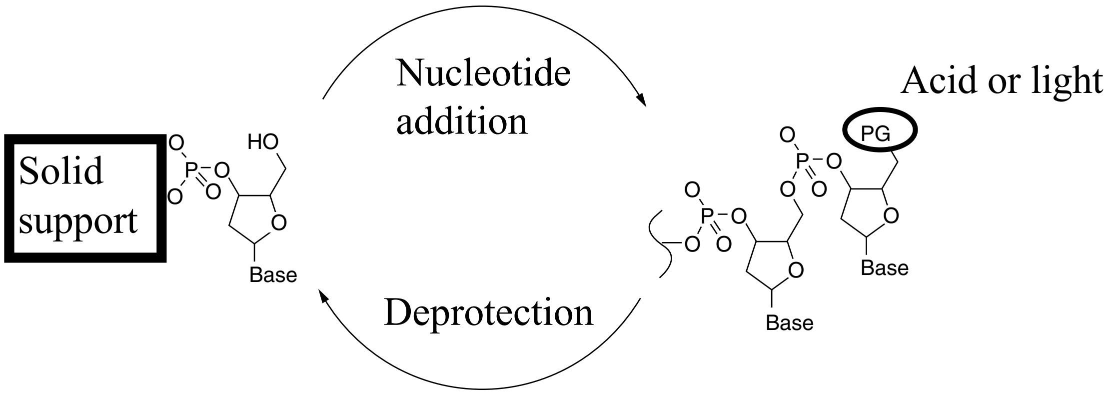


Figure 1.3: In-situ synthesis of oligonucleotides. The oligonucleotides are built on the glass array one base at a time. At each step, the base is added via the reaction between the hydroxyl group 5′ of the terminal base and the phosphate group of the next base. There is a protective group on the 5′ of the base being added, which prevents the addition of more than one base at each step. Following addition, there is a deprotection step at which the protective group is converted to a hydroxyl group to allow addition of the next base.

liquid (Figure 1.1d). This enables higher throughput production of arrays because each probe can be spotted on several arrays without the need to return the pins to the sample plates.

The typical printing process follows five steps (Figure 1.2):

1. The pins are dipped into the wells to collect the first batch of DNA.
2. This DNA is spotted onto a number of different arrays, depending on the number of arrays being made and the amount of liquid the pins can hold.
3. The pins are washed to remove any residual solution and ensure no contamination of the next sample.
4. The pins are dipped into the next set of wells.
5. Return to step 2 and repeat until the array is complete.

In the final phase of array production, the surface of the array can be *fixed* so that no further DNA can attach to it. There are many fixing processes that depend on the precise chemistry on the surface of the glass. The desired outcome is always the same: we do not want DNA target from the sample to stick to the glass of the array during hybridisation, so the surface must be modified so this does not happen. It is also common to modify the surface so that the glass becomes more hydrophilic because this aids mixing of the target solution during the hybridisation stage. Some microarray production facilities do not fix their arrays.

In-Situ Synthesised Oligonucleotide Arrays

These arrays are fundamentally different from spotted arrays: instead of presynthesising oligonucleotides, oligos are built up base-by-base on the surface of the array (Figure 1.3). This takes place by covalent reaction between the 5′ hydroxyl group of

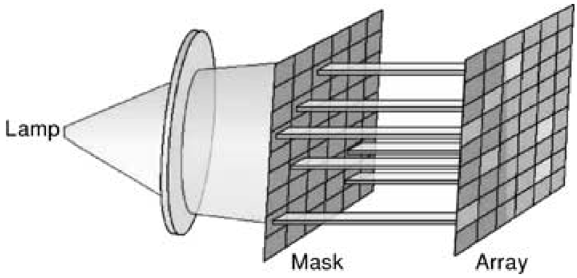


Figure 1.4: Affymetrix technology. Affymetrix arrays are manufactured using in-situ synthesis with a light-mediated deprotection step. During each round of synthesis, a single base is added to appropriate parts of the array. A mask is used to direct light to the appropriate regions of the array so that the base is added to the correct features. Each step requires a different mask. The masks are expensive to produce, but once made, it is straightforward to use them to manufacture a large number of identical arrays. (Reproduced with Permission from Affymetrix Inc.)

the sugar of the last nucleotide to be attached and the phosphate group of the next nucleotide. Each nucleotide added to the oligonucleotide on the glass has a protective group on its 5′ position to prevent the addition of more than one base during each round of synthesis. The protective group is then converted to a hydroxyl group either with acid or with light before the next round of synthesis. The different methods for deprotection lead to the three main technologies for making in-situ synthesised arrays:

1. Photodeprotection using masks: this is the basis of the Affymetrix® technology.
2. Photodeprotection without masks: this is the method used by Nimblegen and Febit.
3. Chemical deprotection with synthesis via inkjet technology: this is the method used by Rosetta, Agilent and Oxford Gene Technology.

Affymetrix Technology

Affymetrix arrays use light to convert the protective group on the terminal nucleotide into a hydroxyl group to which further bases can be added. The light is directed to appropriate features using *masks* that allow light to pass to some areas of the array but not to others (Figure 1.4). This technique is known as photolithography and was first applied to the manufacture of silicon chips. Each step of synthesis requires a different mask, and each mask is expensive to produce. However, once a mask set has been designed and made, it is straightforward to produce a large number of identical arrays. Thus Affymetrix technology is well suited for making large numbers of "standard" arrays that can be widely used throughout the community.

Maskless Photodeprotection Technology

This technology is similar to Affymetrix technology in that light is used to convert the protective group at each step of synthesis. However, instead of using masks, the light is directed via micromirror arrays, such as those made by Texas Instruments. These are solid-state silicon devices that are at the core of some data projectors: an array of mirrors is computer controlled and can be used to direct light to appropriate parts

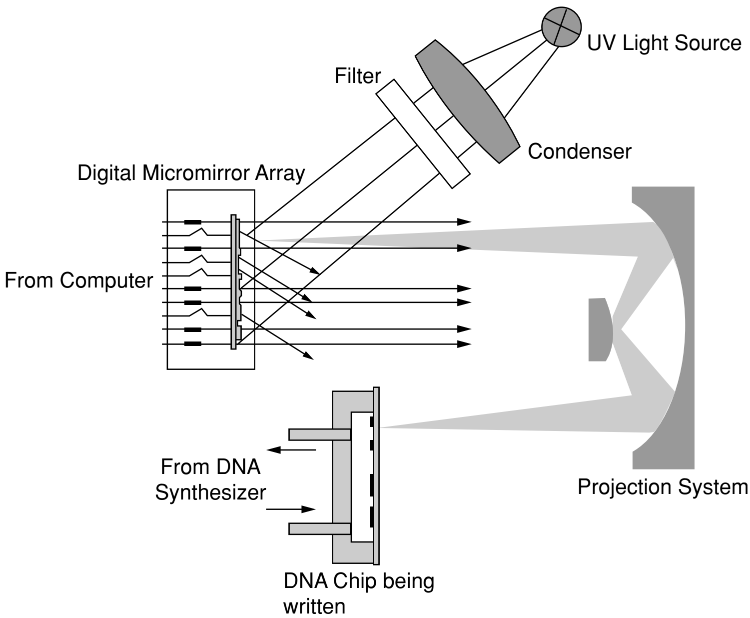


Figure 1.5: Maskless photodeprotection. This system also uses light-mediated deprotection. However, instead of using a physical mask, the array is synthesised using a computer-controlled micromirror array. This consists of a large number of mirrors embedded on a silicon chip, each of which can move between two positions: one position to reflect light, and the other to block light. At each step, the mirrors direct light to the appropriate parts of the array. This technology is used by Nimblegen and Febit.

of the glass slide at each step of oligonucelotide synthesis (Figure 1.5). This is the technology used by Nimblegen and Febit.

Inkjet Array Synthesis

Instead of using light to convert the protective group, deprotection takes place chemically, using the same chemistry as a standard DNA synthesiser. At each step of synthesis, droplets of the appropriate base are fired onto the desired spot on the glass slide via the same nozzles that are used for inkjet printers; but instead of firing cyan, magenta, yellow and black ink, the nozzles fire A, C, G and T nucleotides (Figure 1.6).

One of the main advantages of micromirror and inkjet technologies over both Affymetrix technology and spotted arrays is that the oligonucleotide being synthesised on each feature is entirely controlled by the computer input given to the array-maker at the time of array production. Therefore, these technologies are highly flexible, with each array able to contain any oligonucleotide the operator wishes. However, these technologies are also less efficient for making large numbers of identical arrays.

Synthesis Yields

The different methods of oligonucleotide synthesis have different coupling efficiencies: this is the proportion of nucleotides that are successfully added at each step of synthesis. Photodeprotection has a coupling efficiency of approximately 95%, whereas

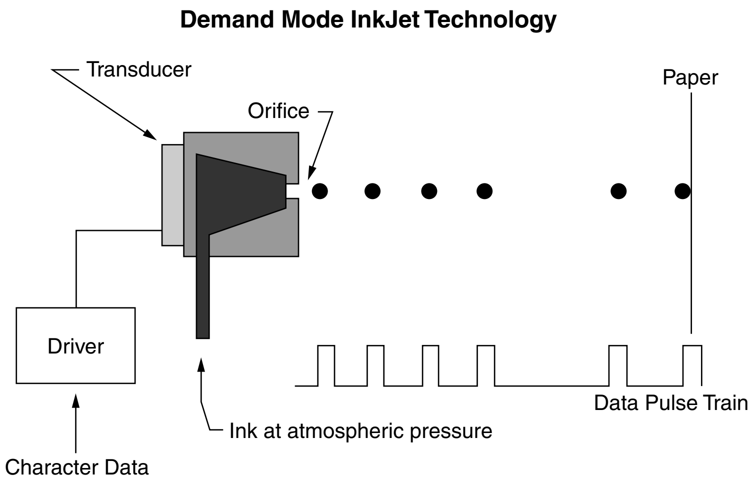


Figure 1.6: Inkjet array synthesis. This technology uses chemical deprotection to synthesise the oligonucleotides. The bases are fired onto the array using modified inkjet nozzles, which, instead of firing different coloured ink, fire different nucleotides. At each step, the appropriate nucleotide is fired onto each spot on the array. The nozzles are computer controlled, so any oligonucleotides can be synthesised on the array simply by specifying the sequences in a computer file. This is the technology used by Rosetta, Agilent and Oxford Gene Technology.

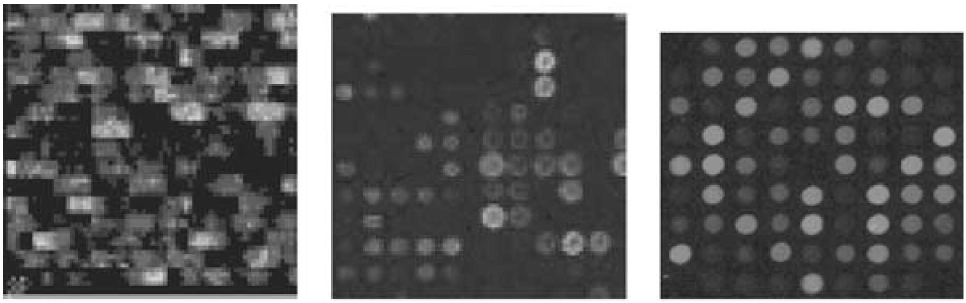
acid-mediated deprotection of dimethoxytrityl protecting groups has a coupling efficiency of approximately 98%. The effect on the yield of full-length oligos is dependent on the length of the oligonucleotide being synthesised: the longer the oligonucleotide, the worse the yield. This dependence is multiplicative, so that even a small difference in coupling efficiency can make a large difference in the yield of long oligonucleotides (Table 1.2).

The composition of the final population of oligonucleotides produced depends on whether or not a capping reaction is included during synthesis. Capping is used by Affymetrix and prevents further synthesis on a failed oligonucleotide. As a result,

TABLE 1.2

Oligonucleotide Length (*s*)	**Coupling Efficiency (*p*)**	**Oligonucleotide Yield (p^s)**
25	95%	28%
25	98%	60%
60	95%	5%
60	98%	30%

Note: The yield of in-situ synthesised oligonucleotides of desired length s depends on the coupling efficiency p according to the formula yield $= p^s$. So the longer the oligonucleotide, the worse the yield. Photodeprotection has a coupling efficiency of approximately 95%, while chemical deprotection has a coupling efficiency of approximately 98%. For a 25-base oligonucleotide, the yields are 28 and 60%, respectively. For a 60-base oligonucleotide, the yields are 5 and 30%, respectively. This is why Affymetrix is restricted to making 25-base oligonucleotides: the coupling efficiency is too low to produce longer oligos. Companies using chemical deprotection are able to synthesise 60-base oligos with similar yield to Affymetrix's 25-base oligos.



(a) (b) (c)

Figure 1.7: Array quality. (a) On Affymetrix arrays the features are rectangular regions. The masks refract light, so there is leakage of signal from one feature to the next. The Affymetrix image-processing software compensates for this by using only the interior portions of the features. **(b)** Spotted arrays produce spots of variable size and quality. This image shows some of this variation; we cover image processing of spotted arrays in detail in Chapter 4. **(c)** Inkjet arrays tend to be of the highest quality, with regular, even spots. (*Please see also the color section at the middle of the book.*)

all oligonucleotides on a feature will have the same start, but will be of different lengths (e.g., with a coupling efficiency of 95%, each feature will be 4.8% monomers, 4.5% dimers, 4.3% trimers, etc.). In contrast, uncapped oligonucleotides allow further synthesis to take place. Therefore, all the oligonucleotides on a feature will be of similar length but may contain random deletions (e.g., with a coupling efficiency of 95% and synthesis of 20 mers, the average probe length would be 19 bases, with such probes containing one deletion).

Spot Quality

The quality of the features depends on the method of array production (Figure 1.7). Spotted array images can be of variable quality, and Chapter 4 is dedicated to the bioinformatics of image processing associated with these arrays. Affymetrix arrays have the problem that the masks refract light, so light leaks into overlapping features; Affymetrix compensates for this with their image-processing software, so the user need not worry about this problem. Inkjet arrays tend to produce the highest quality features.

SECTION 1.3 USING MICROARRAYS

There are four laboratory steps in using a microarray to measure gene expression in a sample (Figure 1.8):

1. Sample preparation and labelling
2. Hybridisation
3. Washing
4. Image acquisition

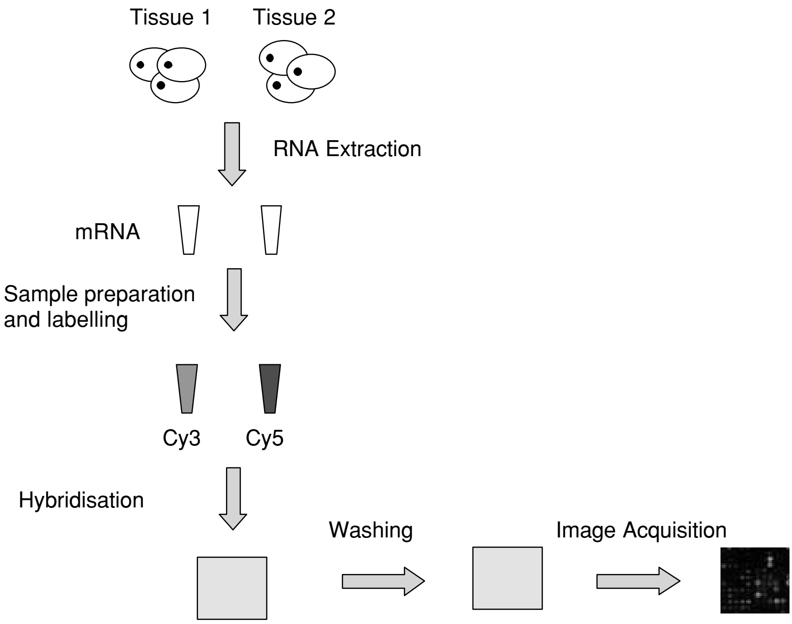


Figure 1.8: Steps in using a microarray. The first step is to extract the RNA from the tissue(s) of interest. With most technologies, it is common to prepare two samples and label them with two different dyes, usually Cy3 (green) and Cy5 (red). The samples are hybridised to the array simultaneously and incubated for between 12 and 24 hours at between 45 and 65°C. The array is then washed to remove sample that is not hybridised to the features.

Sample Preparation and Labelling

There are a number of different ways in which a sample can be prepared and labelled for microarray experiments. In all cases, the first step is to extract the RNA from the tissue of interest. This procedure can be difficult to reproduce, and there is much variability among the individual scientists performing the extraction.

The labelling step depends on the technology used. For the Affymetrix platform, one constructs a biotin-labelled complementary RNA for hybridising to the GeneChip®. The protocols are very carefully defined by Affymetrix,[4] so every Affymetrix laboratory should be performing identical steps. This has the advantage that it is easier to compare the results of experiments performed in different Affymetrix laboratories, because the procedures they will have followed should be the same.

Although it is possible to hybridise complementary RNA to other types of microarrays, it is much more common to hybridise a complementary DNA to these arrays. In the past, the DNA has been radioactively labelled, but now most laboratories use fluorescent labelling, with the two dyes Cy3 (excited by a green laser) and Cy5 (excited by a red laser). In the most common experiments, two samples are hybridised to the arrays, one labelled with each dye; this allows the simultaneous measurement of both

[4] See the reference to the Affymetrix manual at the end of the chapter.

samples. In the future, it is possible that more than two labelled samples could be used.

There are three common ways to make labelled cDNA. The most common method is direct incorporation by reverse transcriptase. The mRNA is primed with a poly-T primer: this starts the reverse transcription from the polyadenylation signal at the 3′ untranslated region (UTR) of the mRNA. In addition to the nucleic acids added for the transcription reaction (dA, dC, dG and dT), a proportion of dCTP (or sometimes dUTP used in place of dT) to which the fluorescent Cy dye has been covalently attached is added to the solution. This means that a proportion of the "C"s in the cDNA product have Cy fluorophors attached to them.

The transcripts produced by this method can be between 0 and 3,000 bases long, and are typically a few hundred bases. They will always be complementary to the 3′ end of the mRNA, so oligonucleotide or cDNA probes on the array must be in the 3′ region of the mRNA, otherwise they will not detect the labelled target produced. Fortunately, the 3′ UTR also tends to be the most variable region of genes, so this is an aid in designing specific probes (Chapter 3).

The next most common method is indirect labelling. This method also uses a reverse transcription reaction, primed from the 3′ end of the mRNA with a poly-T primer. However, instead of using fluorescently labelled dC, the reaction takes place with an amino-allyl-modified dC. This is a much smaller molecule than the Cy-modified dCTP, so the reverse transcription is more efficient. Following reverse transcription, the cDNA is reacted with an active ester of the dye, so that the dye becomes attached to the modified dCs in the cDNA. This method also has the advantage that each target has the same "foreign" base incorporated at the same rate. This contrasts with direct incorporation, where Cy5 is incorporated less well by reverse transcriptase than Cy3.

As with direct incorporation, indirect labelling produces transcripts of a few hundred bases, complementary to the 3′ end of the mRNA. This has similar implications for the design of probes for the array.

The third and least common method for labelling is by random primed labelling using the Klenow fragment of DNA polymerase I. The first step is a reverse transcription reaction, which generates a single-stranded cDNA. The cDNA is then primed with random primers and extended using the Klenow fragment of DNA polymerase I in the presence of labelled dC. The product is a mixture of shorter labelled transcripts, complementary to both strands of the gene.

Because the labelled fragments are on both strands, there is greater potential for cross-hybridisation, and so it is important to check for cross-hybridisation on both strands when designing probes for the arrays (Section 3.3).

Hybridisation

Hybridisation is the step in which the DNA probes on the glass and the labelled DNA (or RNA) target form heteroduplexes via Watson–Crick base-pairing (Figure 1.9). Hybridisation is a complex process that is not fully understood. It is affected by many

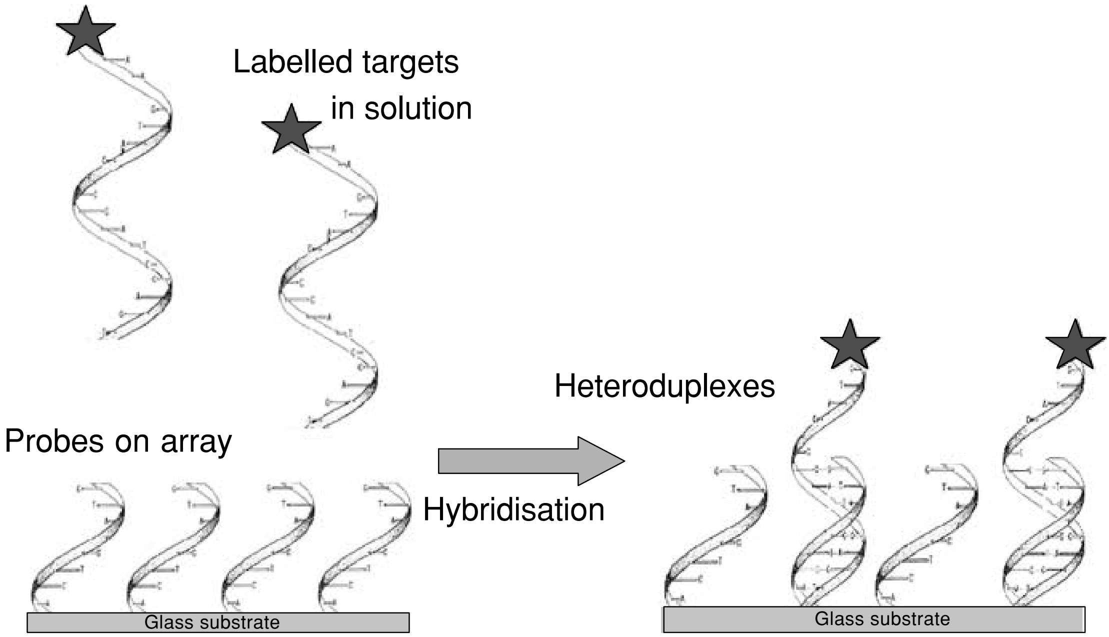
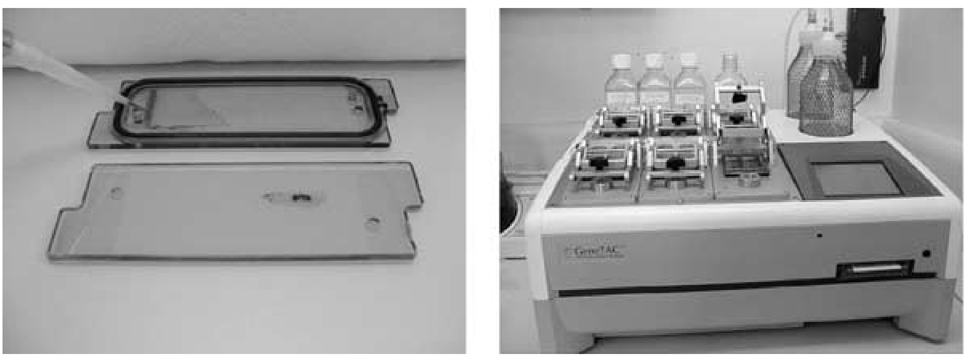


Figure 1.9: Hybridisation. In the hybridisation process, labelled target in solution forms heteroduplexes with probes on the array via Watson–Crick base-pairing between the probes and the target. Unbound target is then washed off the array, so that the only fluorescent signal on the array is in the heteroduplexes. The microarray measures the level of fluorescence on each of the features, and from this we infer the absolute or relative amount of DNA bound to each feature on the array.

conditions, including temperature, humidity, salt concentrations, formamide concentration, volume of target solution and operator.

There are two main methods for hybridisation: manual and robotic. In a manual hybridisation, the array is placed in a hybridisation chamber (Figure 1.10a). The scientist



(a) (b)

Figure 1.10: Hybridisation systems. (a) Manual hybridisation takes place in chamber; the target is injected onto the slide; the chamber is sealed with a cover slip and placed in an incubator. **(b)** Hybridisation station; this station has six chambers in which all hybridisation, incubation and washing takes place. Robotic hybridisation reduces the variability of microarray experiments, but care must be taken not to break the arrays in the robot. This GeneTac station is located at the Mouse Genetics Unit in Harwell, Oxfordshire.

injects the hybridisation solution containing the target onto the array under a cover slip before sealing the chamber. The chamber is placed in an incubator which keeps the array at the correct temperature; some incubators also agitate the array to ensure mixing of the hybridisation solution over the surface of the array. Most hybridisations take place over a period of 12 to 24 hours.

Alternatively, hybridisation can be performed robotically by a hybridisation station (Figure 1.10b). Robotic hybridisation has the advantage over manual hybridisation in that it provides much better control of the temperature of the target and slide. The consistent use of a hybridisation station also reduces the variability between hybridisations and operators.

Most hybridisations are performed at between 45 and 65°C, depending on the type of array used. With oligonucleotide arrays, arrays with different-length oligonucleotides may require different temperatures. The addition of formamide enables mixing of the hybridisation solution of the target over the array but has the effect of decreasing the apparent melting temperature of duplexes. This has the positive benefit of reducing spatial hybridisation irregularities on the array – a matter that is discussed in Chapters 4 and 5. Different laboratories have used a wide range of formamide concentration in the target solution, between 0 and 50%. The use of no formamide at 65°C gives approximately equivalent thermodynamic conditions as the use of 50% formamide at 45°C.

It is also usual to include Na^+ in the hybridisation solution.[5] The less Na^+ present, the greater the stringency of the hybridisation; the thermodynamic effect of Na^+ concentration is well characterised and described in Chapter 3. Most hybridisations take place in approximately 1M Na^+ [between 3 and 5 standard saline citrate (SSC)].

It is also common to add DNA to the hybridisation solution that blocks unwanted cross-hybridisation. The two most common additions are some type of repetitive DNA that masks genomic repeat sequence, such as COT-1, and either poly-A or poly-T to mask the polyadenylation sites on the cDNA.

Washing

After hybridization, the slides are washed. There are two reasons for this. The first is to remove excess hybridisation solution from the array. This ensures that the only labelled target on the array is the target that has specifically bound to the features on the array and thus represents the DNA that we are trying to measure.

The second reason is to increase the stringency of the experiment by reducing cross-hybridisation. This can be achieved either by washing in a low-salt wash (e.g., 0.1 SSC and 0.1 SDS, a detergent that removes grease) or with a high-temperature wash. In either case, the aim is that only the DNA complementary to each of the features will remain bound to the features on the array. Most automatic hybridisation stations include a washing cycle as part of the automated process.

5 It is also possible, but much less common, to use K^+.

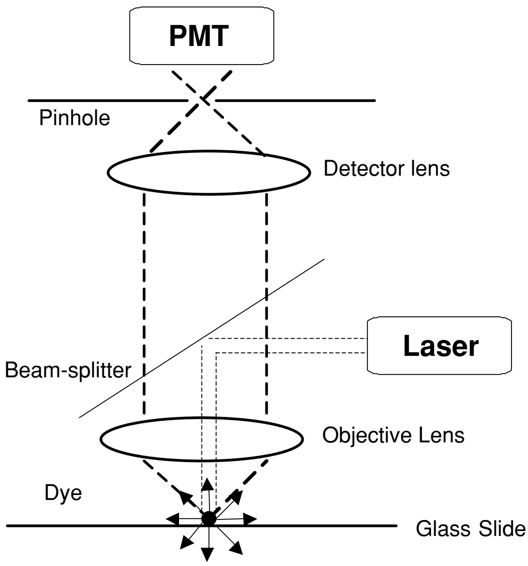


Figure 1.11: Workings of a scanner. The majority of microarray scanners work in a similar way. A laser is used to excite the dyes incorporated into the heteroduplexes on the surface of the array; the fluorescence of the dye is then measured by a PMT and converted to digital signal. Each pixel on the scanned image represents a single point of measurement of fluorescence from the laser. The slide (or in the case of some scanners, the optics) is moved so that the laser excites the whole of the slide. Two-colour arrays are scanned twice: once with a green laser (for Cy3; excitation wavelength is 550 nm and emission wavelength is 581 nm) and once with a red laser (for Cy5; excitation wavelength is 649 nm and emission wavelength is 670 nm).

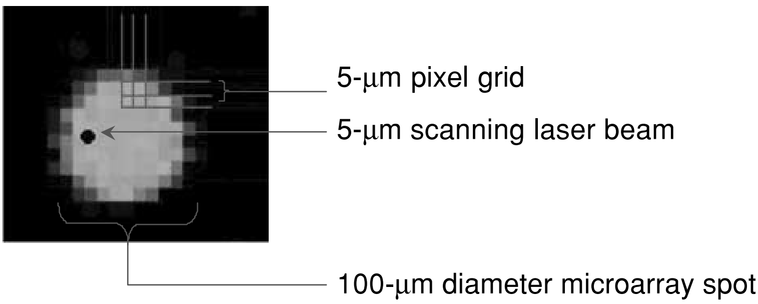
Image Acquisition

The final step of the laboratory process is to produce an image of the surface of the hybridised array. The heteroduplexes on the array, where the target has bound to the probe, contain dye that fluoresces when excited by light of an appropriate wavelength. The slide is placed in a **scanner**, which is a device that reads the surface of the slide. Most scanners have a similar design (Figure 1.11). The scanner contains one or more lasers that are focussed onto the array: most scanners for two-colour arrays use two lasers.

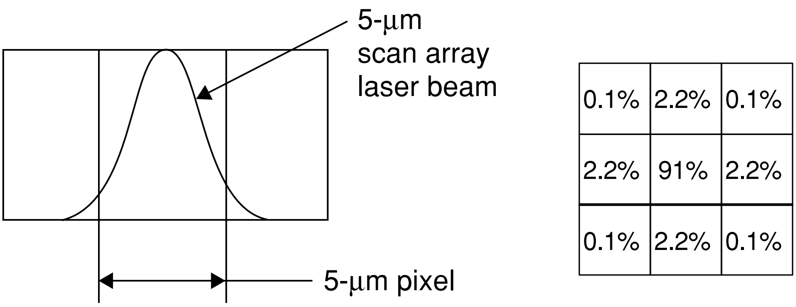
Each pixel on the digital image represents the intensity of fluorescence induced by focussing the laser at that point on the arrray (Figure 1.12a). The dye at that point will be excited by the laser and will fluoresce; this fluorescence is detected by a photomultiplier tube (PMT) in the scanner. In order to scan the whole array, the laser must be focussed on every point on the array. This is achieved either by moving the slide so that the laser can focus on different points, or by shifting the optics to achieve the same result.

It is usual for the size of the physical space represented by the pixel to be the same as the spot size of the laser. When this is the case, the majority of the light measured at that pixel comes from that point on the array (Figure 1.12b). It is also possible to set the pixel size to be smaller than the laser spot. In that case, much of the light at each pixel comes from neighbouring areas on the array (Figure 1.12c). This has the effect of blurring the image and can mask minor irregularities in the feature (Figures 1.12d and 1.12e).

With two-colour arrays, the output of the scanner is usually two monochrome images: one for each of the two lasers in the scanner (Figure 1.13a). These are combined to create the familiar red–green false colour images of microarrays. Both the monochrome and two-colour images are usually stored as tagged image file format (TIFF). The array data is stored in 16 bits. This means that the intensity of each pixel in each channel is quantified as a 16-bit number, which takes values between 0 and $2^{16} - 1$, which is equal to 65,535. Since background is approximately 100, and saturation can occur when the average pixel intensity is larger than 50,000, the microarray can detect intensities over an approximately 500-fold dynamic range. With 10-μm pixel sizes, a typical microarray image will be 7,500 × 2,200 pixels. This means that each of the two



(a)



(b)

Figure 1.12: The pixels comprising a feature. (a) A false-colour image of the pixels from a single scan of a 100-μm microarray feature. The size of the laser spot is 5 μm. The pixel size has been set to 5 μm so that each pixel represents the area from the size of the laser spot. **(b)** The intensity of light from a laser is normally distributed. With a 5-μm laser size and 5-μm pixel size, 91% of the emission from the array resulting from the laser is measured at that pixel. **(c)** With a 10-μm laser size and 5-μm pixel size, the majority of light from the laser is measured in neighbouring pixels. This has the effect of blurring the image. **(d)** Two neighbouring features on an array with a streak through them, measured with a laser spot size of 5 μm and a pixel size of 5 μm. The streak is clear on both spots and so the spot can be identified as problematic. **(e)** The same features scanned with laser spot size of 10 μm and a pixel size of 5 μm. The streak has become blurred. (*Please see also the color section at the middle of the book.*)

(*continued*)

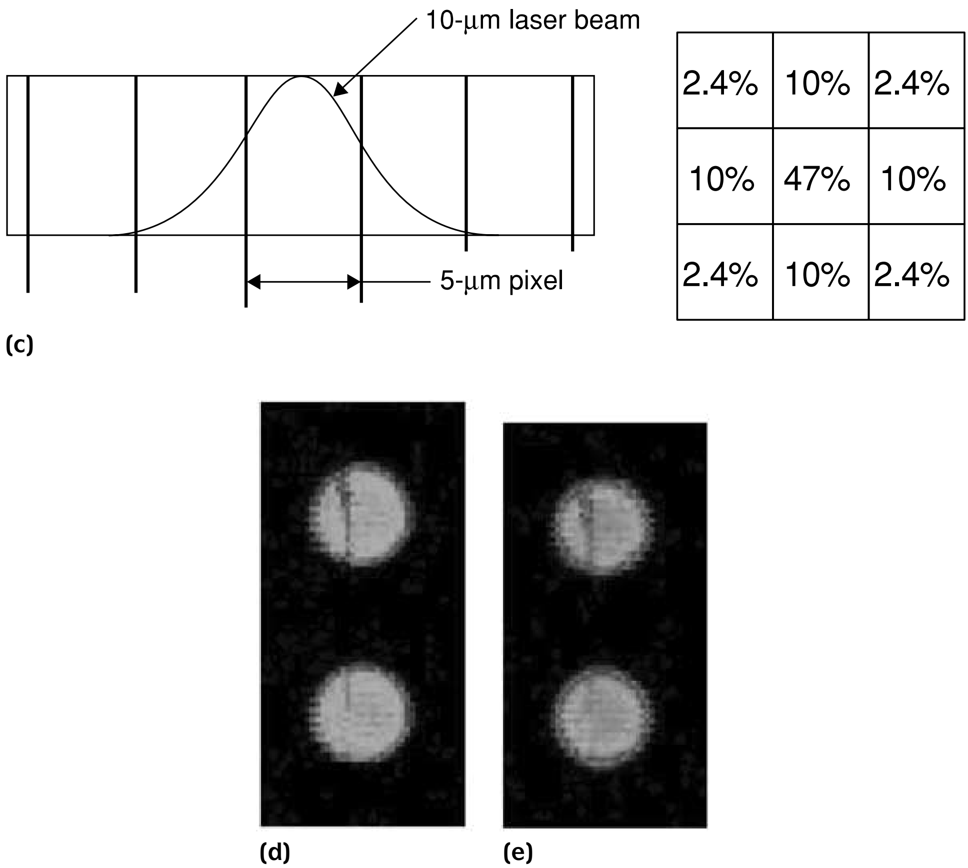


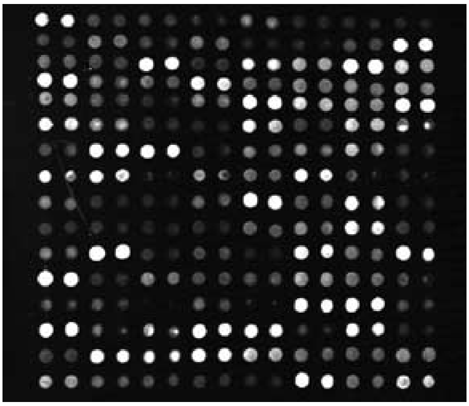
Figure 1.12: (*continued*)

TIFF images is 32 Mb; these are large files so if you are producing a large number of microarray images, data storage becomes an important consideration (Chapter 11).

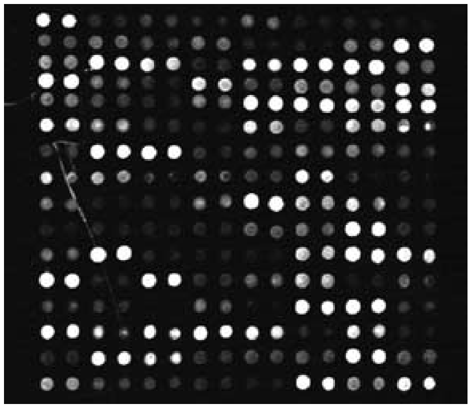
The pixel resolution of the image should be chosen so that each feature has sufficient pixels to make the measurement of the intensity of the feature robust from pixel-to-pixel noise. It is normally recommended that there should be at least 50 pixels per feature on the array.

KEY POINTS SUMMARY

- Making and using microarrays is a complex laboratory process.
- There are many sources of variability in microarray experiments.
- The main microarray technologies are
  - Spotted cDNAs, the most common type of microarray;
  - Spotted oligonucleotides: increasingly common and better quality than spotted cDNAs;



(a)



(b)

(c)

Figure 1.13: Output of scanners. (a) This is the scanner output for a part of a microarray – in this case one of twelve 16 × 16 blocks of features. This is the monochrome image of the Cy3 (green) channel. **(b)** The scanner output for the same part of the array but using the Cy5 (red) channel. **(c)** It is usual to combine the two monochrome images into a composite false-colour image of the array. Green features correspond to features that are expressed more in the sample labelled with Cy3 than the sample labelled with Cy5, and so will be bright in (a) and dark in (b). Similarly, red spots will be bright in (b) and dark in (a). Yellow features have a similar level of expression in both samples. Dark features are low expressed in both samples. (*Please see also the color section at the middle of the book.*)

- Light-directed in-situ synthesised arrays: e.g., Affymetrix, the most common commercial platform; and
- Inkjet in-situ synthesised arrays, the highest quality arrays but not widely available.

■ The steps in using a microarray are

- Target preparation, which introduces most variability;
- Hybridisation, also a source of variability, which can be reduced by robotics;
- Washing, which can increase stringency – variability can be reduced by robotics; and